The not-so-Irish spurge: Euphorbia hyberna (Euphorbiaceae) and the Littletonian plant 'steeplechase'

The not-so-Irish spurge: *Euphorbia hyberna* (Euphorbiaceae) and the Littletonian plant "steeplechase"

GEMMA E. BEATTY[1,2], JACK J. LENNON[1,2], CHRIS J. O'SULLIVAN[1] and JIM PROVAN[1,2*]

[1] *School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL*

[2] *Institute for Global Food Security, Queen's University Belfast*

Running title: Phylogeography of *Euphorbia hyberna*

[*]Corresponding author: Dr Jim Provan
School of Biological Sciences,
Queen's University Belfast,
97 Lisburn Road,
Belfast BT9 7BL
E-mail: J.Provan@qub.ac.uk
Tel: +44 (0)28 9097 2280
Fax: +44 (0)28 9097 5588

The disjunct distributions of the Lusitanian flora, which are found only in southwest Ireland and northern Iberia, and are generally absent from intervening regions, have been of great interest to biogeographers. There has been a long debate as to whether Irish populations represent relicts that survived the Last Glacial Maximum (LGM; *ca.* 21ka), or whether they recolonized from southern refugia following the retreat of the ice, and if so, whether this was directly, due to long distance dispersal, or successively, in the manner of a "steeplechase", with the English Channel and Irish Sea representing successive "water-jumps" that have to be successfully crossed. In the present study, we used a combined palaeodistribution modelling and phylogeographical approach to elucidate the glacial history of the Irish spurge, *Euphorbia hyberna*, the sole member of the Lusitanian flora that is also thought to occur naturally in southwestern England. Our findings suggest that the species persisted through the LGM in several southern refugia, and that northern populations are the result of successive recolonization of Britain and Ireland during the postglacial Littletonian warm stage, akin to the "steeplechase" hypothesis.

ADDITIONAL KEYWORDS**:** *Euphorbia hyberna* – Irish spurge – Last Glacial Maximum – Lusitanian flora – palaeodistribution modelling – phylogeography

INTRODUCTION

The present-day distributions of many species are the result of climatic fluctuations during the Quaternary period (*ca.* 2.5 MYA – present; Webb & Bartlein, 1992; Comes & Kadereit, 1998; Hewitt, 2003; Provan & Bennett, 2008). During the most recent (Weichselian) glacial period in Europe (*ca.* 115 KYA – 20 KYA), these species persisted in refugia associated with the southern peninsulas of Iberia, Italy and the Balkans, although fossil and genetic data also indicate survival in more northerly, or "cryptic", refugia (Taberlet *et al.,* 1998; Hewitt, 1999; Stewart & Lister, 2001; Bennett & Provan, 2008; Provan & Bennett, 2008). Phylogeographic analyses in particular have allowed the reconstruction of various patterns of postglacial recolonization from one or more of these refugia that have led to current-day species distributions (Taberlet *et al.,* 1998; Petit *et al.,* 2003).

Ireland represents a particularly interesting biogeographical case-study in postglacial recolonization. As an island on the western fringe of the European continent, it has been isolated from Britain by postglacial sea-rise for *ca.* 15 KY, twice as long as Britain itself has been isolated from mainland Europe (Edwards & Brooks, 2008). Ireland has a relatively impoverished flora, having only *ca.* 800 species compared to nearly 1,200 in Britain and 3,500 in France, and with only 18 of these not found in Britain compared to 375 British species absent from Ireland (Reid, 1913; Webb, 1983). This has been attributed to the recolonization of Britain and Ireland during the postglacial Littletonian warm stage being somewhat akin to a "steeplechase", with the English Channel and Irish Sea representing successive "water-jumps" that have to be successfully crossed (Mitchell & Ryan 1992; Jones, 2011).

The so-called "Lusitanian" element of the Irish flora comprises a number of species that exhibit a disjunct distribution, being found in southern and western Ireland, as well as in

43 northern Spain, but mostly absent from intervening countries (Matthews, 1926; Praeger,
44 1933, 1939; Baker, 1959; Webb, 1983). The Lusitanian distribution has been the subject of
45 debate for many years, with some botanists claiming it to be the result of persistence in
46 separate Irish and Iberian refugia during the last glaciation (Forbes, 1846; Praeger, 1933),
47 whilst others favoured long-distance dispersal from a southern refugium, claiming that full
48 glacial conditions precluded *in situ* survival in Ireland (Reid, 1913). Recently, the first
49 phylogeographic studies on three plant species exhibiting Lusitanian distributions, *Daboecia*
50 *cantabrica*, *Pinguicula grandiflora* and *Saxifraga spathularis*, concur with the latter
51 hypothesis and suggest that they achieved their current distributions as a result of
52 recolonization from a range of refugia in Iberia and the Bay of Biscay following the last
53 glacial maximum (LGM; *ca.* 21 ka; Beatty & Provan, 2013, 2014).
54 Irish spurge (*Euphorbia hyberna*) is the sole member of the Lusitanian flora that also
55 occurs naturally, although very sporadically, in southwestern England (Devon and Cornwall)
56 and in central and southern France, as well as in northern Spain and southwestern Ireland
57 (Figure 1a). Unlike the previously studied Lusitanian plant species named above, which have
58 minute, dust-like seeds conducive to long-distance dispersal, *E. hyberna* has large (3-5 mm)
59 seeds with far less capacity for dispersal. Consequently, given its occurrence in the
60 intervening countries between Ireland and Spain, we employed a combined palaeodistribution
61 modelling and phylogeographical approach to test whether the species persisted during the
62 LGM in northern refugia, or whether the colonization of Ireland could have taken place in the
63 sequential fashion of the Littletonian plant "steeplechase". We analysed the distribution of
64 genotypes at one chloroplast and one nuclear marker from samples across the species' range
65 in combination with the palaeodistribution model to identify the locations of glacial refugia
66 during the LGM, and to elucidate how postglacial recolonization has resulted in the species'
67 current-day distribution.

MATERIALS AND METHODS

STUDY SPECIES

Euphorbia hyberna is a monoecious perennial which can be found growing in woodland glades, hedgerows and on shaded stream banks, and is most easily recognized from April to June when it flowers. The yellow-green flowers are formed on tall stems which can be up to 60 cm tall, and lack petals and sepals. The sap is poisonous, containing phorbyl esters, and can cause irritation to the skin. In England, the species is classed by the IUCN as "Vulnerable", based on data from The Vascular Plant Red Data List (IUCN, 2001).

SAMPLING AND DNA EXTRACTION

Samples of *E. hyberna* were collected from six locations across the species' Irish range in Counties Cork and Kerry. Leaf samples were also obtained from herbarium specimens from a further 19 locations in Ireland, three locations in England, eight locations from France, and from 36 locations spanning the species' entire Spanish distribution (see Fig.1 and Supporting Information Table S1 for details of locations and numbers of samples). DNA was extracted from field-collected material using a modified CTAB (cetyl trimethyl ammonium bromide) protocol (Doyle & Doyle, 1987) and from herbarium samples using Qiagen DNeasy kits.

PALAEODISTRIBUTION MODELLING

We modelled suitable climate envelopes for *E. hyberna* at the LGM (*ca*. 21 ka) in order to predict potential refugia, using an ensemble modelling (EM) approach (R Biomod2 package; Thuiller *et al.*, 2012). Species occurrence data (412 spatially unique records; Fig.2a) were obtained from the Global Biodiversity Information Facility data portal (http://www.gbif.org/). Ten different distribution models using as explanatory covariates the first four principle

93 component (PC) scores from 19 BIOCLIM variables (WorldClim data set 1950–2000;
94 Hijmans *et al.*, 2005) were generated at 2.5 arc-minute resolution extracted from an area
95 bounded by 15°W to 30°E and 35°N to 60°N. These first four principal components
96 accounted for 99% of the variation in the climate data. We used this approach because using
97 the BIOCLIM variables directly failed even when selecting from the 19 variables by
98 removing the most strongly correlated ones (r>0.7). While the current-climate models looked
99 plausible and had good AUC (area under the curve) of the ROC (Receiver-Operator
100 Characteristic) scores (AUC>0.8), the resultant palaeoclimatic projections were very different
101 from each other; this was traced to instability caused by the strong cross correlations still
102 present between the explanatory variables (Dormann *et al*., 2013). The ten models were
103 screened for performance using the conventional 70/30 training/validation data partition.
104 Each model was run ten times using this partition. Eight models passed this filter (FDA,
105 flexible discriminant analysis, did not converge and we excluded SRE, surface range
106 envelope, because it had a consistently lower AUC goodness-of-fit measure (Supporting
107 Information Table S2). These remaining eight models (ten replicates of each) were then
108 combined using the ROC measure to give an EM, using the median measure, and propagating
109 the uncertainty from the training/test split from the 80 fitted models. The rationale for an EM
110 approach is that such a composite model often outperforms individual models (Seni & Elder,
111 2010), though it is by no means settled how best this can be done for species distribution
112 modelling. This EM was then supplied with reconstructed LGM data [Community Climate
113 System Model (CCSM); Palaeoclimate Modelling Intercomparison Project Phase II:
114 http://pmip2.lsce.ipsl.fr/] to identify potential *E. hyberna* refugial areas. We calculated a
115 multivariate environmental similarity surface (MESS; Elith *et al.*, 2010) to ensure that the
116 species range projected at the LGM climate was not outside the current climate – in other
117 words, that the LGM projections were not extrapolations outside current climate space.

CHLOROPLAST *TRN*S-*TRN*G SEQUENCING

In total, 200 samples were sequenced for the chloroplast *trnp*S–*trn*G intergenic spacer. A product was initially amplified and sequenced using the universal *trn*S-*trn*G primers described in Zhang *et al*., (2005), and a pair of species-specific primers were subsequently designed: Eh-trnS 5′-CATCTCTCCCGATTGAAAAGG-3′ and Eh-trnG 5′-TAAACTATACCCGCTACGATACAA-3′. For herbarium samples from which the complete product could not be amplified in a single polymerase chain reaction (PCR), the region was amplified as two or three overlapping fragments using the primers described in Supporting Information Table S3. PCR was carried out on a MWG Primus thermal cycler (Ebersberg, Germany) using the following parameters: initial denaturation at 94 °C for 3 min followed by 45 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, extension at 72 °C for 1 min and a final extension at 72 °C for 5 min. PCR was carried out in a total volume of 20 μL containing 200 ng genomic DNA, 10 pmol of each primer, 1× PCR reaction buffer, 200 μM each dNTP, 2.5 mM $MgCl_2$ and 0.5 U GoTaq Flexi DNA polymerase (Promega, Sunnyvale CA). 5 μL volumes of PCR products were resolved on 1.5% agarose gels and visualized by ethidium bromide staining, and the remaining 15 μL were ExoSAP-purified and sequenced in both directions using the BigDye sequencing kit (v3.1; Applied Biosystems, Foster City, CA) and run on an AB 3730XL DNA analyser (Life Technologies; Carlsbad, CA). Sequence lengths ranged from 548 bp – 563 bp, and the overall alignment was 628 bp in length.

SINGLE-COPY NUCLEAR DNA (SCNDNA) SEQUENCING

Primers to amplify an anonymous single-copy nuclear DNA locus (Eh-E04) were developed using the ISSR cloning method described in Beatty, Philipp & Provan (2010). The 214 bp region was amplified in 200 individuals using the following primers: Eh-E04-F

143 5′-TTCCAAATTCCAATTCTGTGC-3′ and Eh-E04-R
144 5′-CATCATCATTCAATTAACAAATAAA-3. PCR and sequencing were carried out as
145 described above.
146
147 PHYLOGEOGRAPHIC ANALYSIS
148 For the scnDNA locus, potential recombination was assessed using the Hudson & Kaplan
149 (1985) test in the DnaSP software package (V5.10; Librado & Rozas 2009). As no evidence
150 of recombination was detected, haplotypes were resolved for individuals exhibiting two or
151 more heterozygous positions using the PHASE program (V2.1; Stephens & Donnelly 2003)
152 implemented in DnaSP. DNA sequences were aligned in BIOEDIT (V7.0.9.0; Hall, 1999).
153 Median-joining networks for both regions were constructed using the NETWORK software
154 package (V4.5.1.6; www.fluxus-engineering.com). Any reticulations in the networks were
155 broken following the rules described in Pfenninger & Posada (2002).
156 To identify groups of populations in Spain and France associated with potential refugial
157 areas, we performed a spatial analysis of molecular variance (SAMOVA) using the software
158 package SAMOVA (V1.0; Doupanloup, Schneider& Excoffier, 2002) for both of the data
159 sets. This program uses a simulated annealing approach based on genetic and geographical
160 data to identify groups of related populations. The program was run for 10,000 iterations for
161 K = 2 to 10 groups, from 200 initial conditions, and the most likely structure was identified
162 using the maximum value of Φ_{CT}, the proportion of genetic variation between groups of
163 populations, that did not include any groups of a single population.
164 Levels of haplotype diversity (H) and nucleotide diversity (π) at both the chloroplast and
165 scnDNA locus were calculated for mainland European, English and Irish samples using
166 DnaSP. To account for differences in sample sizes, effective numbers of haplotypes (h_e)
167 were also calculated using HAPLOTYPE ANALYSIS 1.05 (Eliades & Eliades, 2009).

RESULTS

PALAEODISTRIBUTION MODELLING

The current climate species distribution ensemble model (Fig.2) and its component eight distribution models based on current climate (Supporting Information Fig. S1) have what is usually considered to be high AUC values (Supporting Information Table S2; EM AUC=0.974) and, as is often the case, good visual congruence between current distribution records and areas of suitable climate (except the individual MARS model). Example model projections for the component eight models for the LGM are given in Supporting Information Fig. S2, and the climatic data (PC scores) used are mapped in Supporting Information Figs. S3 and S4. Notable in Fig.2 is the general spread of climatically suitable areas beyond the current *Euphorbia hyberna* range, such as north western Ireland, western Britain and Belgium. There are also a few *Eurphorbia hyberna* records in areas of apparently low current climatic suitability, such as northern Italy and Slovenia. In contrast to the current EM, the paleodistribution EM is spatially rather more coherent in that there is a wide band of suitable climate extending from northern Spain and southwest France across the Bay of Biscay and up to the south and west of Ireland and the LGM ice sheet (Fig. 2). Within this range of suitable past climate there are areas identified by the EM as particularly suitable and so more likely as potential refugia. These include in Spain the four main areas of Castile and León, northern Galicia, northern Aragon and Cataluña, and in France a small area on the island of Corsica, a large area centered on the Poitou Charentes region, and a smaller one in the Maritime Alps. Finally, there are three suitable regions in the area now covered by the Atlantic: in the Bay of Biscay at the western tip of Brittany, larger area to the south of this, and an isolated area well to the south of Ireland. The multivariate environmental similarity surface (MESS; Supporting Information Fig. S5) indicated that only a small amount of

193 extrapolation into novel climate space occurred, and that this was primarily in the north
194 where the ice sheet was present.
195
196 PHYLOGEOGRAPHIC ANALYSIS
197 18 and 14 haplotypes were identified for the chloroplast *trn*S–*trn*G intergenic spacer and the
198 anonymous Eh-E04 single copy nuclear locus, respectively (Figs. 3a/b; GenBank accession
199 numbers XXX – XXX and XXX – XXX). The SAMOVA analyses based on the scnDNA
200 locus did not identify more than a single group comprised of more than one individual, but
201 the analysis based on the chloroplast *trn*S-*trn*G region identified $K = 4$ groups ($\Phi_{CT} = 0.728$;
202 Fig. 1). Some degree of geographical structuring of haplotypes was evident for both loci, but
203 particularly for the chloroplast *trn*S–*trn*G intergenic spacer (Fig. 3a). The dark blue, light
204 blue, pink and white haplotypes were restricted to the eastern part of the main continental
205 distribution of *E. hyberna*, around the Pyrenees, whilst the yellow, light yellow, brown and
206 purple haplotypes were only found west of this region. This was broadly reflected in the
207 distributions of the haplotypes displayed by the anonymous Eh-E04 single copy nuclear locus
208 (Fig. 3b). For both loci, the distribution of the green haplotype was of particular note. This
209 haplotype was found in England, along with the ubiquitous red haplotype, but in mainland
210 Europe was restricted to a single location in the extreme northeast of the Basque Country in
211 Spain, adjacent to the Bay of Biscay. All diversity statistics for the two markers indicate a
212 decrease in levels of genetic diversity from mainland Europe, through England, to Ireland,
213 where both loci were monomorphic (Table 1).

DISCUSSION

The EM for the current climate and distribution (Fig. 2) suggests that *Euphorbia hyberna* is not climatically limited with respect to future expansion into much of Britain and Ireland. Belgium is another notably climatically suitable but unoccupied area. These unoccupied but apparently suitable areas serve as a reminder that apparently suitable locations need not be occupied, since species have requirements and traits other than climate that must be present for effective recolonization and population persistence. *E. hyberna* has large, globose seeds, with no obvious modifications to aid dispersal, and absence from climatically suitable regions may occur due to low capacity for dispersal, or simply because its other niche requirements, such as suitable substrate and/or nutrients, are not met. Alternatively, such areas may be attributable to the modelling approach, despite the high AUC scores. In general, it is becoming apparent that model uncertainty is not well dealt with by current species distribution models, including those used here, particularly uncertainty caused by structural model misspecification (e.g. ignoring autocorrelation in explanatory covariates), response variable quality (e.g. variation in spatial recording effort), and low-quality predictive covariate data or predictive covariate data that represents an extrapolation (Beale & Lennon, 2012). For this reason the paleodistribution model results here, and elsewhere, are best taken as a broadly indicative rather than as definitive of potential refugia locations, despite the high AUC score (Beale, Lennon & Gimona, 2008).

The paleodistribution model identified a number of potential refugia at the LGM, several of which overlap the current distribution of *E. hyberna.* These areas are comparable in size and climatic suitability (by definition) to the current distribution. The ensemble model is of nominally good fit but of course there is no way to establish from modelling alone if any particular putative refugium was occupied at the LGM; several may have been unoccupied as

239 a consequence of unsatisfied non-climatic niche requirements. However, based on a
240 combination of this palaeodistribution modelling and the spatial structuring observed in
241 genetic data from two independent loci (particularly the chloroplast data) it appears that the
242 species persisted in multiple southern refugia during the glaciations. The findings of the
243 present study further highlight the complex nature of refugial persistence in the Iberian
244 Peninsula. The previous concept that each of the three southern peninsulas of Iberia, Italy
245 and the Balkans represented a single refugial area has now been superseded by the idea of
246 "refugia-within-refugia", originally proposed for Iberia, but now with several examples from
247 the other peninsulas (Gomez and Lunt, 2007). The potential persistence of *E. hyberna* in
248 separate eastern and western refugia, as suggested by the genetic data and to some extent the
249 palaeodistribution model, which indicated an area of suitable habitat in Galicia, mirrors the
250 patterns found for *D. cantabrica* (Beatty & Provan, 2013). The existence of two SAMOVA
251 groups (shown in blue and pink in Fig. 1) apparently associated with the Pyrenees could be
252 explained by divergence of these groups in separate microrefugia / microhabitats within, or in
253 close proximity to, this region. A previous phylogeographical study on white oaks (*Quercus*
254 spp.) in Iberia also identified two lineages (designated Lineage A and Lineage C) restricted to
255 the eastern Pyrenees that were hypothesized to have originated in separate Cataluñan refugia
256 (Olalde *et al.*, 2002). *E. hyberna* is most commonly found in oak woodland, and it is possible
257 that the two taxa, *E. hyberna* and *Quercus* spp., could have persisted together during the
258 LGM. The general west / east distribution of the pink and blue groups respectively along the
259 Pyrenees, could indicate persistence in separate microrefugia in the Pyrenees. Such refugia
260 could be difficult to identify using the palaeodistribution model at the resolution used,
261 particularly given that mountainous regions can provide a range of spatial and elevational
262 habitats that could conceivably promote divergence of lineages (Bennett & Provan, 2008;
263 Médail & Diadema, 2009; Holderegger & Thiel-Egenter, 2009; Stewart *et al.*, 2010).

Our findings strongly suggest that the present disjunct distribution of *E. hyberna* has resulted from recolonization of England and Ireland from a southern refugium, possibly in the region of the Bay of Biscay (see below). A similar scenario for the postglacial recolonization of Ireland was found in the two previous phylogeographic studies on plants exhibiting Lusitanian distributions (Beatty & Provan, 2013, 2014). In general, Irish populations exhibited lower levels of genetic diversity than those in Spain, with *E. hyberna* completely lacking any private haplotypes, classic signatures of postglacial recolonization rather than glacial persistence (Provan & Bennett, 2008). Thus, although the palaeodistribution model indicated possible suitable habitat for *E. hyberna* along the edge of the continental shelf as far as the limits of the British-Irish ice sheet at the LGM, the phylogeographic evidence indicates that if such a refugium did exist on land that is now submerged, and was the source of the northern populations, it must have been situated much further south.

The fact that *E. hyberna* is the only plant with a Lusitanian distribution that also occurs naturally in England means that alternative theories on the postglacial recolonization of Lusitanian species can be examined. In the previous phylogeographic studies on elements of the Lusitanian flora, the extremely disjunct distribution of present-day populations meant that it was not possible to differentiate between recolonization as a result of long-distance dispersal, and a more gradual "stepping-stone" pattern of recolonization followed by subsequent extirpation of the species from intervening areas (Beatty & Provan, 2013, 2014). The sequential decrease in genetic diversity observed in continental, English and Irish populations of *E. hyberna* in both markers studied is entirely consistent with the "steeplechase" scenario following the end of the Weichselian glaciation (Matthews, 1926; Mitchell & Ryan 1992; Jones, 2011). At both loci, the haplotypes indicated in green, which in Spain are restricted to populations in the extreme north, close to the French border and adjacent to the Bay of Biscay, are found along with the most common red haplotype in

England, but only the latter is present in Ireland. Our evidence that *E. hyberna* colonized Ireland in the fashion of the Littletonian plant "steeplechase" thus raises the intriguing possibility that other Lusitanian species might have achieved their present-day distributions in a similar manner, with no need to invoke extreme long-distance dispersal events. A similar scenario was originally proposed nearly 100 years ago based on biogeographical data (Stapf, 1914, 1916), but the present study represents the first test of this hypothesis at the intraspecific level (although see Valtueña, Preston & Kadereit, 2012 for a species with a similarly disjunct distribution, despite not being strictly a Lusitanian one).

There also remains the possibility that the present disjunct distribution of *E. hyberna* (and other members of the Lusitanian flora) might be due to anthropogenic transport. The extremely low levels of genetic variation in Irish populations could be consistent with a single, recent introduction, and there are many documented cases of the introduction, deliberately or accidentally, of mammals into Ireland by humans since Mesolithic times (reviewed in Montgomery *et al.* 2014). The role of humans in the introduction of the Lusitanian flora, however, has rarely been considered. With the exception of the Strawberry tree, *Arbutus unedo*, which has been planted as an ornamental, botanists generally now believe that anthropogenic introduction of these species to Ireland is unlikely, and that natural postglacial recolonization is the best explanation for their present-day distributions (Sealy 1949). Although no macrofossil or palynological evidence exists for *E. hyberna* in Ireland, at least one other member of the Lusitanian flora, *Daboecia cantabrica*, is known to have been present as far back as the Gortian (Holsteinian) interglacial (*ca.* 428-302 ka; Woodell 1958; Coxon 1996).

In conclusion, our findings suggest that the present-day disjunct distribution of *Euphorbia hyberna* did not result from glacial survival in separate northern and Iberian refugia, contrary to Forbes' (1846) original idea on the origin of the Lusitanian distribution, but instead is due

to the recolonization of England and Ireland from a southern refugium, possibly in the area around the Bay of Biscay as also previously indicated for *Daboecia cantabrica*. Whilst these results cast doubt on one early biogeographical hypothesis on the origin of the Lusitanian flora, they do provide support for another, namely Stapf's (1914, 1916) theory of progressive dispersal in the fashion of the Littletonian "steeplechase". Furthermore, they highlight the complex processes responsible for the present-day distribution of genetic variation in the Iberian Peninsula, as well as those operating at species' rear-edges in general, where populations often represent reservoirs of unique genetic diversity (Hampe & Petit, 2005; Provan & Maggs 2012).

ACKNOWLEDGEMENTS

We are extremely grateful to Robert Beatty and Javier Amigo Vasquez (Universidad de Santiago de Compostela Herbario) for assistance with field sampling. Valuable herbarium specimens were provided by Colin Kelleher and Noeleen Smyth (National Botanic Gardens of Ireland, Glasnevin), Sarah Whild (University of Birmingham Herbarium), Carmen Fernández-Carvajal (Universidad de Oviedo Herbario), Elena de Paz Canuria (Universidad de León Herbario), Iñaki Aizpuru and Anaïs Mitxelena (Alto do Zorroaga s.n. Herbario), Javier Hernández (Universidad de Salamanca Herbario), Enriqueta Martín-Consuegra Fernández (Universidad de Córdoba Herbario) and Laurent Gautier (University of Geneva Herbarium). We are grateful to Valerie Hall, Ian Montgomery and Keith Bennett for many interesting and helpful discussions on the glacial history of Ireland and its flora and fauna, and to two anonymous referees for their helpful comments on the original version of the manuscript. This research was supported by Research Grant RP-126 from the Leverhulme Trust.

338 REFERENCES

339

340 **Baker HG. 1959.** The contribution of autecological and genecological studies to our knowledge of the past
341 migrations of plants. *The American Naturalist*, **93**: 255-272.

342 **Beale CM, Lennon JJ. 2012.** Incorporating uncertainty in predictive species distribution modelling.
343 *Philosophical Transactions of the Royal Society B Series*, **367**: 247-258.

344 **Beale CM, Lennon JJ, Gimona A. 2008.** Opening the climate envelope reveals no macroscale associations
345 with climate in European birds. *Proceedings of the National Academy of Sciences of the United States of*
346 *America*, **105**: 14908-14912.

347 **Beatty GE, Philipp M, Provan J. 2010.** Unidirectional hybridization at a species' range boundary:
348 implications for habitat tracking. *Diversity and Distributions*, **16**: 1-9.

349 **Beatty GE, Provan J. 2013.** Postglacial dispersal, rather than *in situ* glacial survival, best explains the disjunct
350 distribution of the Lusitanian plant species *Daboecia cantabrica* (Ericaceae). *Journal of Biogeography* **40**:
351 335-344.

352 **Beatty GE, Provan J**. **2014.** Phylogeographical analysis of two cold-tolerant plants with disjunct Lusitanian
353 distributions does not support *in situ* survival during the last glaciations. *Journal of Biogeography* DOI:
354 10.1111/jbi.12371.

355 **Bennett KD, Provan J. 2008.** What do we mean by 'refugia'? *Quaternary Science Reviews*, **27**: 2449-2455.

356 **Cantor SB, Sun CC, Tortolero-Luna G, Richards-Kortum R, Follen M. 1999.** A comparison of C/B ratios
357 from studies using receiver operating characteristic curve analysis. *Journal of Clinical Epidemiology*, **52**:
358 885–892.

359 **Comes HP, Kadereit JW. 1998.** The effect of Quaternary climatic changes on plant distribution and evolution.
360 *Trends in Plant Science*, **3**: 432-438.

361 **Coxon P. 1996.** The Gortian temperate stage. *Quaternary Science Reviews*, **15**: 425–436.

362 **Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B,**
363 **Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell**
364 **D, Lautenbach S. 2013.** Collinearity: a review of methods to deal with it and a simulation study evaluating
365 their performance. *Ecography*, **36**: 27-46.

366 **Doupanloup I, Schneider S, Excoffier L. 2002.** A simulated annealing approach to define the genetic

structure of populations. *Molecular Ecology*, **11**: 2571–2581.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**: 11–15.

Edwards RJ, Brooks AJ. 2008. The Island of Ireland: drowning the myth of an Irish land bridge? *Mind the Gap: Postglacial Colonisation of Ireland*. (ed by Davenport, J.L., Sleeman, D.P. & Woodman, P.C.) pp. 19-34. Irish Naturalists' Journal Special Supplement, Dublin

Eliades N-G, Eliades DG. 2009. *HAPLOTYPE ANALYSIS: software for analysis of haplotype data.* Available at: http:www.uni-goettingen.de/en/134935.html

Elith J, Kearney M, Phillips S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution,* **1**: 330-342.

Forbes E. 1846. On the Connexion between the Distribution of the existing Fauna and Flora of the British Isles, and the Geological Changes which have affected their areas, especially during the epoch of the Northern Drift. *Great Britain Geological Survey Memoir*, **1**: 336-432.

Gómez A, Lunt DH. 2007. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia: evolutionary perspectives on the origins and conservation of European biodiversity* (ed. by S. Weiss and N. Ferrand), pp. 155–188. Springer, Amsterdam.

Hall TA. 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**: 95-98.

Hampe A, Petit RJ. 2005. Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* **8**: 461-467.

Hewitt GM. 1999. Post-glacial recolonisation of European biota. *Biological Journal of the Linnean Society*, **68**: 87-112.

Hewitt GM. 2003. Ice ages: their impact on species distributions and evolution. *Evolution on Planet Earth* (eds. L.J. Rothschild & A.M Lister) pp. 339-361. Academic Press.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Journal of Climatology*, **25**: 1965-1978.

Holderegger R, Thiel-Egenter C. 2009. A discussion of different types of glacial refugia used in mountain biogeography and phylogeography. *Journal of Biogeography*, **36**: 476-480.

Hudson RR, Kaplan N. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 47-164.

IUCN. 2001. *IUCN Red List Categories and Criteria: Version 3.1*. IUCN Species Survival Commission. IUCN, Gland and Cambridge.

Jones RW. 2011. *Applications of Palaeontology: Techniques and Case Studies.* Cambridge University Press, London.

Librado P, Rozas J. 2009. DnaSP V5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**: 1451-1452.

Matthews JR. 1926. The distribution of certain members of the British flora. III. Irish and Anglo-Irish plants. *Annals of Botany*, **40**, 773-797.

Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, **36**: 1333-1345.

Mitchell F, Ryan J. 1992. *Reading the Irish Landscape.* Town House, Dublin.

Montgomery WI, Provan J, McCabe AM, Yalden DW. 2014. Origin of British and Irish mammals: disparate post-glacial recolonization and species introductions. *Quaternary Science Reviews* **98**: 144-165.

Olalde M, Herran A, Espinel S, Goicoechea PG. 2002. White oaks phylogeography in the Iberian Peninsula. *Forest Ecology and Management*, **156**: 89-102.

Petit RJ, Auinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**: 1563-1565.

Pfenninger M, Posada D. 2002. Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): fragmentation, corridor migration and secondary contact. *Evolution*, **56**: 1776–1788.

Praeger RL. 1933. Recent views bearing on the problem of the Irish flora and fauna. *Proceedings of the Royal Irish Academy Section B*, **41**: 125–145.

Praeger RL. 1939. The relations of the flora and fauna of Ireland to those of other countries. *Proceedings of the Linnean Society of London*, **151,** 192–213.

Provan J, Bennett, KD. 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, **23**: 564-571.

Provan J, Maggs CA. 2012. Unique genetic variation at a species' rear edge is under threat from global climate change. *Proceedings of the Royal Society of London Series B* **279**: 39-47.

Reid C. 1913. The relation of the present plant population of the British Isles to the glacial period. *Journal of Ecology* **1,** 42–46.

Sealy JR. 1949. *Arbutus unedo. Journal of Ecology*, **37**: 365-388.

Sejrup HP, Hjelstuen BO, Dahlgren KIT, Haflidason H, Kuijpers A, Nygård A, Praeg D, Stoker MS, Vorren TO. 2005. Pleistocene glacial history of the NW European continental margin. *Marine and Petroleum Geology*, **22**: 1111–1129.

Seni G, Elder JF. 2010. *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*. Morgan & Claypool.

Stapf O. 1914. The southern element in the British flora. *Engler's Botanische Jarbücher*, **50**, 509-525

Stapf O. 1916. A cartographic study of the southern element in the British flora. *Proceedings of the Linnean Society of London*, **129**, 81-92.

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**: 1162-1169.

Stewart JR, Lister AM. 2001, Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, **16**: 608-613.

Stewart JR, Lister AM, Barnes I, Dalén L. 2010. Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society of London Series B*, **277**: 661-671.

Taberlet P, Fumigalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**: 453-464.

Thuiller W, Georges D, Engler R. 2013. BIOMOD 2: Ensemble platform for species distribution modeling. R package version 3.1-25. Available at http://CRAN.R-project.org/package=biomod2

Valtueña FJ, Preston CD, Kadereit JW. 2012. Phylogeography of a Tertiary relict plant, *Meconopsis cambric* (Papaveraceae), implies the existence of northern refugia for a temperate herb. *Molecular Ecology*, **21**: 1423-1437.

Webb DA. 1983. The flora of Ireland in its European context. *Journal of Life Sciences, Royal Dublin Society*, **4**, 143-160.

Webb T, Bartlein PJ. 1992. Global changes during the last 3 million years: climatic controls and biotic response. *Annual Review of Ecology and Systematics*, **23**: 141-173.

Woodell SRJ. 1958. Biological flora of the British Isles L.C. (Ed. 11) No. 266: *Daboecia cantabrica* K. Koch (*D. polifolia* D.Don; *Menziesia polifolia* Juss.). *Journal of Ecology*, **46:** 205–216.

Zhang Q, Chiang TY, George M, Liu JQ, Abbott RJ. 2005. Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Molecular Ecology* **14**: 3513-3524.

Table 1. Diversity statistics by region.

Locus	Mainland Europe					England					Ireland					GenBank
	N	*h*	h_e	*H*	π	*N*	*h*	h_e	*H*	π	*N*	*h*	h_e	*H*	π	Accessions
*trn*S-*trn*G	114	18	3.672	0.782	0.0032	10	2	1.724	0.467	0.0009	76	1	1	-	-	
Eh-E04	234[a]	14	1.377	0.289	0.0015	20[a]	2	1.220	0.189	0.0009	146[a]	1	1	-	-	

[a] Two gene copies sequenced per diploid individual

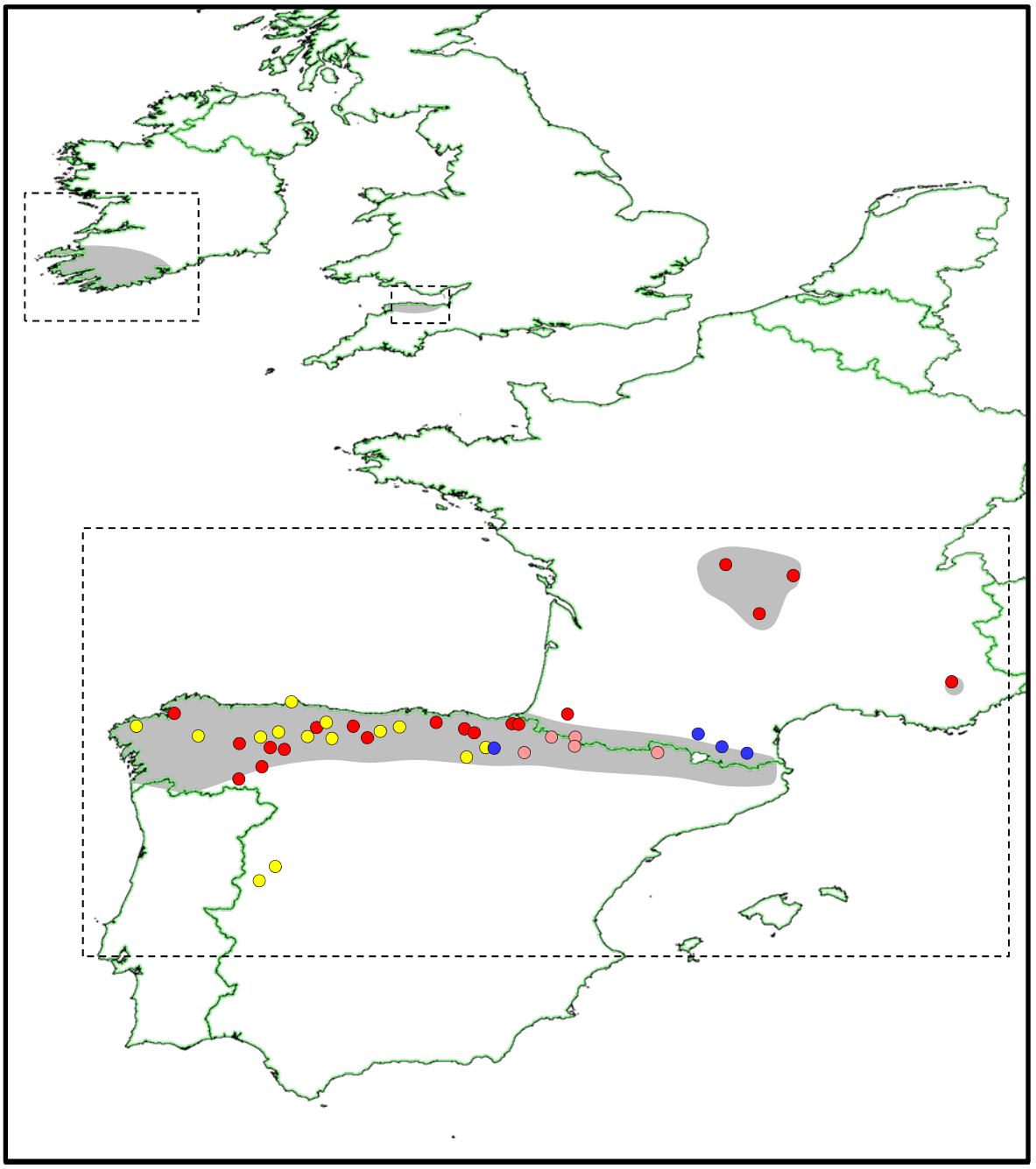
Abbreviations: *N*, number of individuals studied; *h,* number of haplotypes; h_e, effective number of haplotypes*; H*, gene diversity; π, nucleotide diversity.

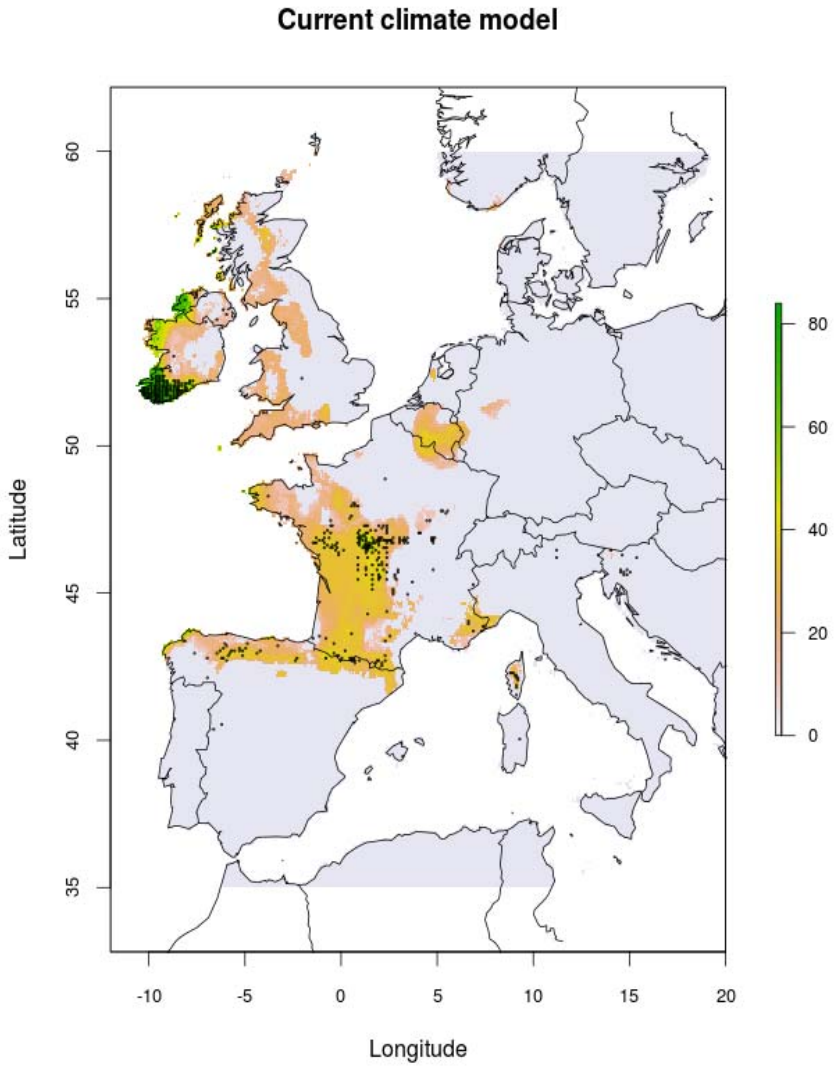
Figure Legends

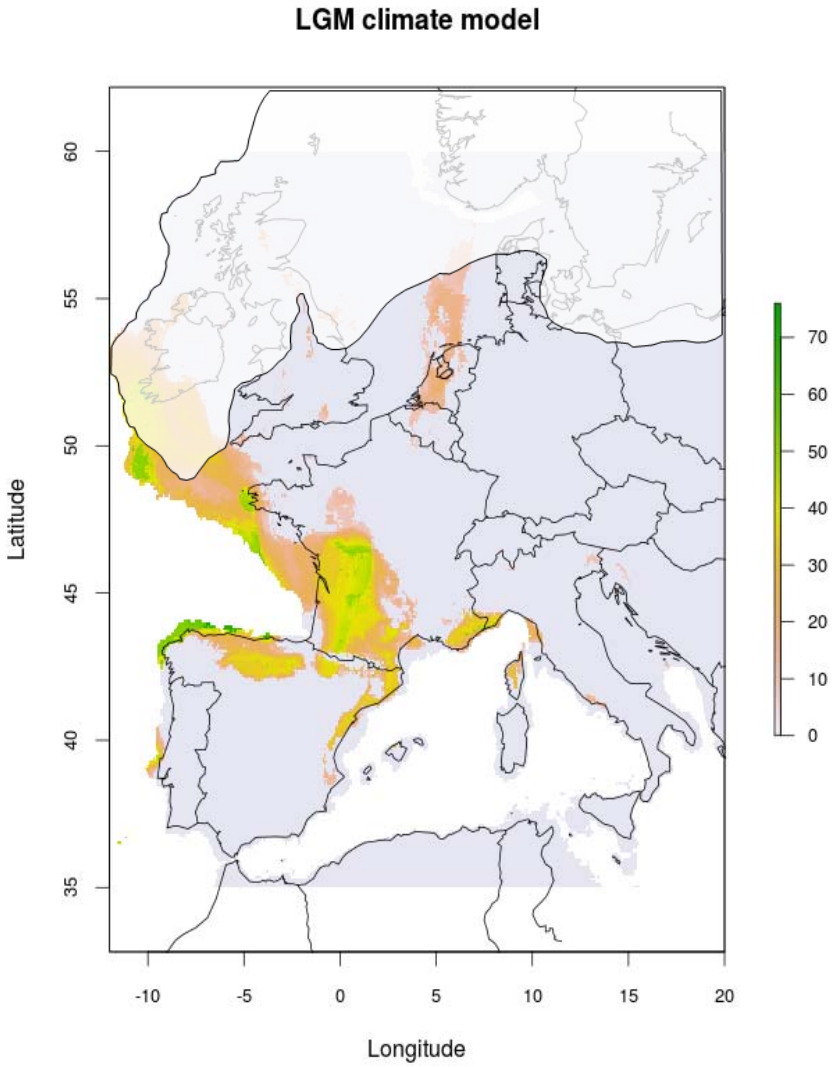
Fig. 1. Present-day distribution of *Euphorbia hyberna* [shaded; based on Webb, 1983, and the Global Biodiversity Information Facility (data.gbif.org)] in Western Europe. Dashed areas show the regions highlighted in Fig. 3a/b. Coloured circles depict assignment of populations to one of K =4 clusters by the SAMOVA analysis.

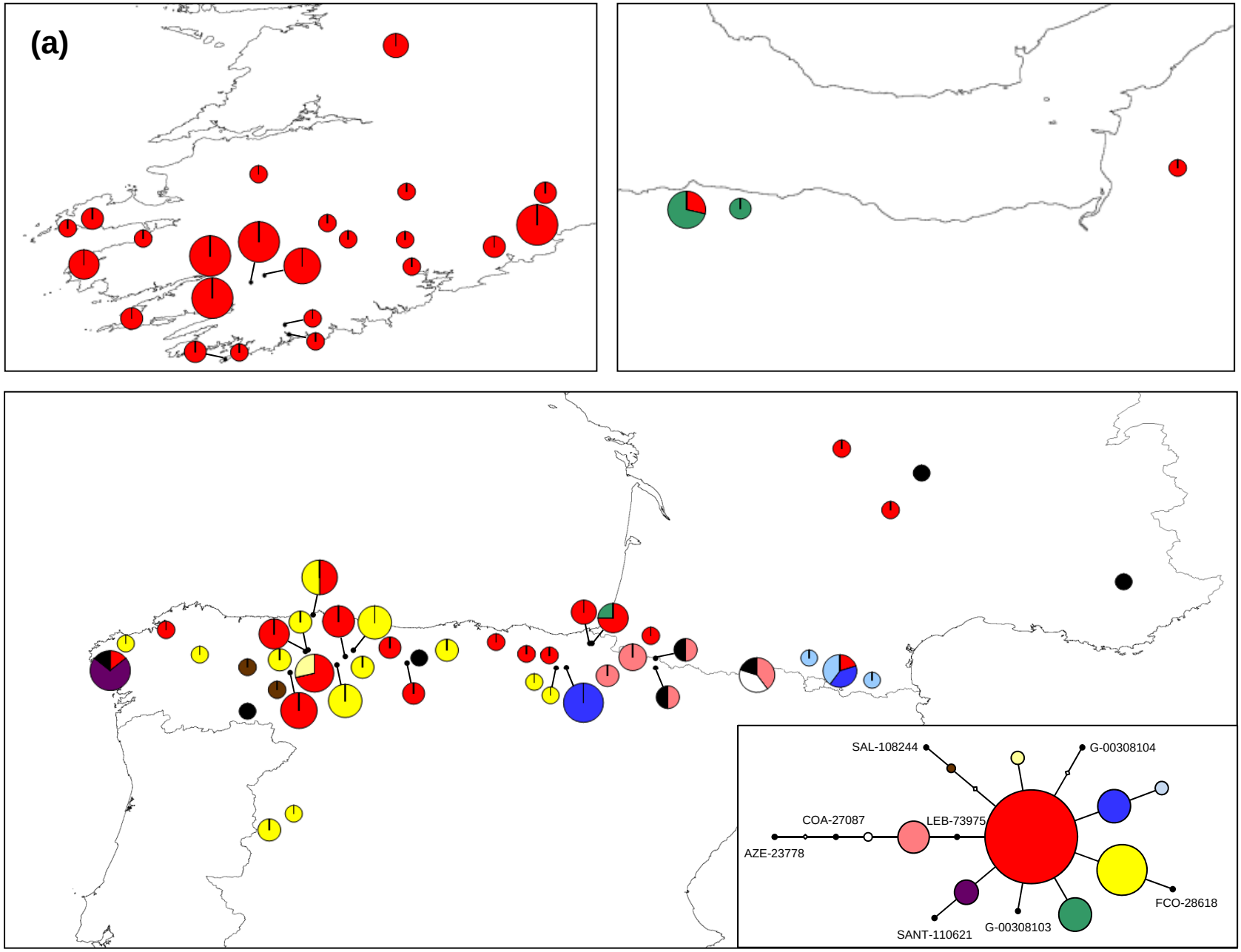
Fig. 2. Modelled distribution of *Euphorbia hyberna* (a) current climate and (b) at the Last Glacial Maximum, (LGM, *ca.* 21 ka). The limits of the ice sheets (after Sejrup *et al.*, 2005) at the LGM are also indicated. Both panels show the modelled species range according to an ensemble model based on ten replicates of eight different model types as the coloured areas. Within this range there is variation in climatic suitability, as indicated by the colour shading.

Fig. 3. Haplotype distributions for (a) chloroplast *trn*S-*trn*G region and (b) nuclear Eh-E04 region for *Euphorbia hyberna* in Western Europe. Pie chart sizes are approximately proportional to sample size, with the smallest circles representing N = 1 (chloroplast locus) or N = 2 (nuclear locus) and the largest representing N = 8 (chloroplast locus) or N = 16 (nuclear locus). In the haplotype networks, circle sizes are approximately proportional to haplotype frequency. Open diamonds represent missing haplotypes and small black circles represent unique haplotypes i.e. those found in a single individual. Codes for single-individual haplotypes refer to IH codes and barcodes given in Supplementary Information Table S1.




Current climate model
Latitude
Longitude
60
55
50
45
40
35
-10
-5
0
5
10
15
20
80
60
40
20
0


LGM climate model
Latitude
Longitude
60
55
50
45
40
35
-10
-5
0
5
10
15
20
70
60
50
40
30
20
10
0


(a)
SAL-108244
G-00308104
COA-27087
LEB-73975
AZE-23778
FCO-28618
SANT-110621
G-00308103

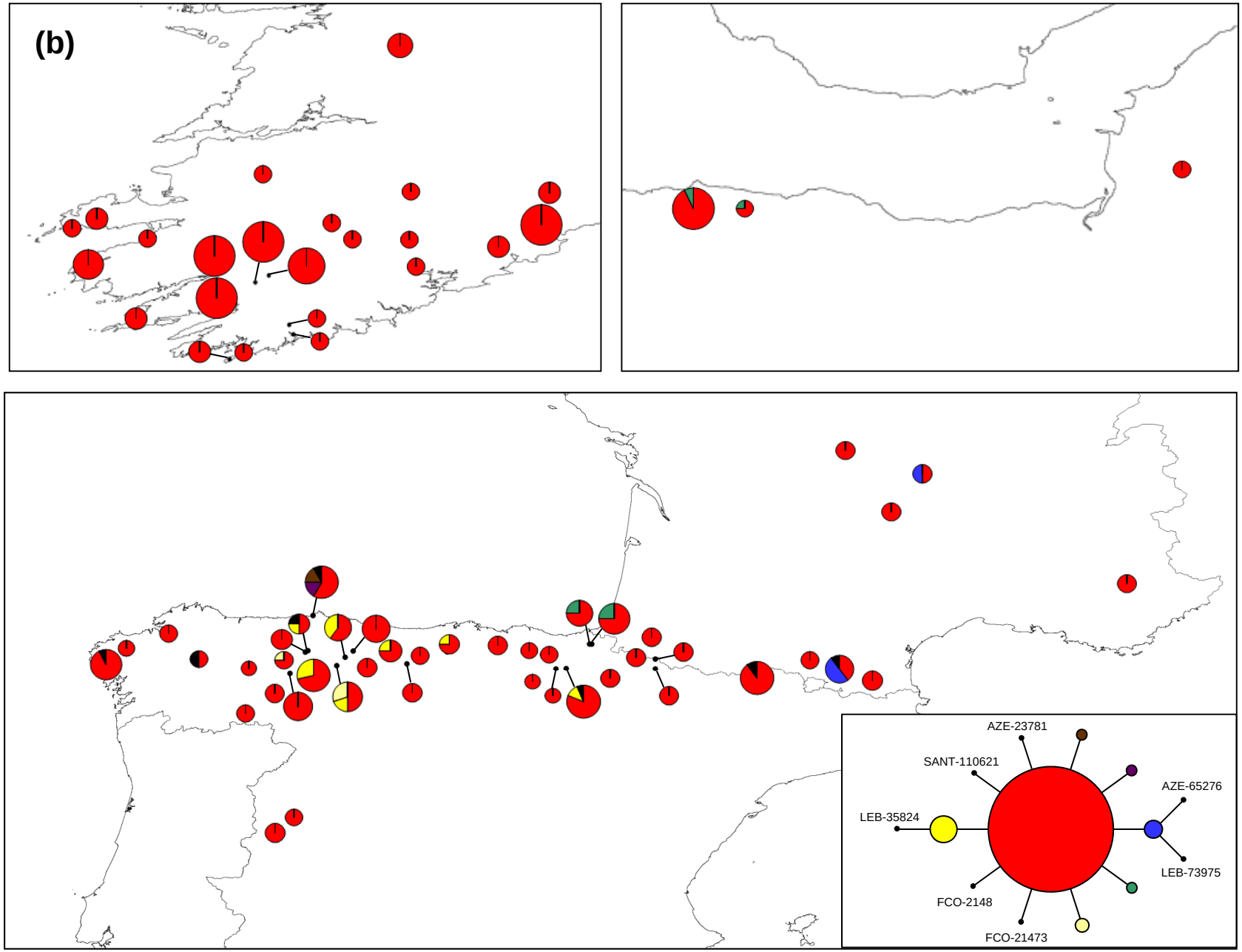

(b)
AZE-23781
SANT-110621
AZE-65276
LEB-35824
LEB-73975
FCO-2148
FCO-21473

Table S1 Populations analysed in this study

Country	Location	Code[a]	Latitude (N)	Longitude (W)	*N*	
					*trn*S-*trn*G	Eh-E04
Ireland	Fahan	DBN-95-1990	52.106499	-10.410695	1	1
	Valentia Island	DBN-87-1980	51.923101	-10.321638	4	4
	Dingle	DBN-53-3790	52.144635	-10.270301	2	2
	Inishturk	BIRM-019143	53.7	-10.1	1	-
	Black Head	DBN-20-1961	51.639034	-10.042036	2	2
	Rossbeigh	DBN-22-1967	52.057273	-9.974015	1	1
	Derrycunihy	IRDE[b]	51.966667	-9.583333	8	8
	Derrycunihy	IRKB[b]	51.971333	-9.581667	8	8
	Glengarriff	IRGG[b]	51.75	-9.566667	8	7
	Cape Clear Island	DBN-11-1964	51.438435	-9.494109	2	2
	Sherkin Island	DBN-55-1975	51.466159	-9.417596	1	1
	Gougane Barra	IRGB[b]	51.835167	-9.345	8	7
	Alleghaun River	DBN-33-1990	52.387281	-9.301819	1	1
	Gurteenowen	IRGU[b]	51.86905	-9.265833	6	6
	Drinagh	DBN-50-1980	51.615221	-9.149051	1	1
	Glendore	DBN-54-1972	51.566787	-9.122982	1	1
	Banteer	DBN-75-1977	52.130557	-8.895895	1	1
	Glennaharee East	DBN-102-1969	52.048487	-8.780994	1	1
	Derrybrien	DBN-18-1992	53.050328	-8.50633	3	3
	Glenville	DBN-48-1988	52.046378	-8.445829	1	1
	Ballyhoura Hills	DBN-104-1969	52.29625	-8.438092	2	2
	Glanmire	DBN-63-1976	51.905429	-8.411502	1	1
	Glendine	DBN-46-1964	52.005661	-7.929738	2	2
	Colligan Wood	IRCW[b]	52.133333	-7.683333	8	8
	Nire Valley	DBN-23-1962	52.277012	-7.634635	2	2

Table S1 (continued)

Country	Location	Code	Latitude (N)	Longitude (W)	*N*	
					*trn*S-*trn*G	Eh-E04
England	Lynmouth	BIRM-019135/37/41/42/44/49/51	51.2	-3.8	7	7
	Badgworthy Water	BIRM-019147	51.2	-3.7	2	2
	Badgworth	BIRM-019148	51.265	-2.873	1	1
France	Pyrénées-Atlantiques	G-00308107	43.250278	-0.879444	1	1
	Ariège	G-00308108	42.933056	1.501111	1	1
	Ax-les-Thermes	AZE-65276	42.76	1.96	5	5
	Aubusson	G-00308109	46.078889	1.992222	1	1
	Pyrénées-Orientales	G-00308110	42.603333	2.448889	1	1
	Chanterelle	G-00308105	45.155556	2.716111	1	1
	Puy de Dome	G-00308104	45.718889	3.186944	1	1
	Alpes-de-Haute-Provence	G-00308103	44.090833	6.232222	1	1
Spain	Mazaricos	LEB-50433/SANT-110621	42.9389	-8.9922	7	7
	Monte Casteso	LEB-31789	43.082449	-8.754298	1	1
	Capela	FCO-18295	43.35	-8.14	1	1
	Outeiro de Rei	FCO-21473	42.98	-7.65	1	1
	Zamora	SAL-108244	42.15	-6.92	1	1
	Leon	COA-27085	42.79	-6.92	1	1
	Caceres	SAL-64349	40.35	-6.59	2	2
	Palacios de Compludo	LEB-101280	42.452437	-6.481113	1	1
	Orallo	LEB-23973	42.963693	-6.425155	2	2
	Igüeña	LEB-78704	42.715476	-6.282805	7	7
	El Cabaco	SAL-90448	40.57	-6.22	1	1
	Puerto de Ventana	LEB-35824/41015	43.044764	-6.053467	4	4
	Oviedo	COA-6411	43.061419	-6.01331	2	2
	Val de Samario	LEB-60837/61455/61451	42.720058	-5.967744	7	7

Table S1 (continued)

Country	Location	Code	Latitude (N)	Longitude (W)	*N*	
					*trn*S-*trn*G	Eh-E04
Spain	Gozón	FCO-02148/LEB-21590	43.59	-5.94	6	6
	Llombera	LEB-54527/54563	42.837471	-5.58183	5	5
	Redipuertas	LEB-16558/17188	42.966384	-5.452345	5	5
	Puebla de Lillo	LEB-37591/44012	43.058921	-5.333235	5	5
	Valdore	LEB-14420/14428	42.791403	-5.200889	2	2
	Llanaves de la Reina	LEB-61452	43.059661	-4.792943	2	2
	Ruesga	LEB-41416	42.865368	-4.530235	2	2
	San Cebrián de Mudá	FCO-28618	42.92	-4.35	1	1
	Zeanuri	AZE-5283	43.048	-3.927	2	2
	Balmaseda	AZE-5284	43.177	-3.203	1	1
	Otxandio	AZE-5282	43.033	-2.65	1	1
	Lagrán	AZE-5280	42.61	-2.628	1	1
	Aretxabaleta	AZE-5285	43.005	-2.491	1	1
	Alava	COA-27906	42.792351	-2.290801	1	1
	Améscoa Baja	AZE-23777/23781	42.796	-2.138	8	8
	Goizueta	AZE-23784	43.162	-1.788	4	4
	Aranaz	AZE-23783	43.162	-1.751	3	3
	Leoz	AZE-23780	42.664	-1.517	2	2
	Ochagavía	AZE-23782	42.964	-1.142	3	6
	Huesca	COA-27087	42.79	-0.8	2	2
	Izaba	AZE-23778	42.94	-0.8	2	2
	Artiga de Lin	LEB-73975/COA-31229	42.681502	0.705355	5	5

[a] IH codes: DBN - National Botanic Gardens of Ireland, Glasnevin; BIRM - University of Birmingham Herbarium; G – University of Geneva Herbarium; SANT - Universidad de Santiago de Compostela Herbario; FCO - Universidad de Oviedo Herbario; LEB - Universidad de León Herbario; COA - Universidad de Córdoba Herbario; SAL - Universidad de Salamanca Herbario; AZE - Alto do Zorroaga s.n. Herbario.
[b] These codes refer to our own, field collected samples

Table S2 AUC values from nine model types fitted using 70/30 training/test split showing ten replicates of each model. SRE was dropped and the remaining eight models used to generate an ensemble model, where the contribution model of each model to the ensemble is weighted by the AUC score.

Model*	Run									
	1	2	3	4	5	6	7	8	9	10
GLM	0.918	0.913	0.907	0.894	0.878	0.916	0.914	0.891	0.924	0.920
GAM	0.944	0.934	0.937	0.929	0.925	0.950	0.934	0.934	0.945	0.951
ANN	0.897	0.893	0.886	0.881	0.871	0.924	0.875	0.868	0.927	0.898
SRE	0.802	0.797	0.792	0.761	0.774	0.785	0.776	0.788	0.796	0.757
CTA	0.942	0.927	0.920	0.895	0.908	0.928	0.916	0.895	0.947	0.927
GBM	0.952	0.939	0.948	0.932	0.934	0.956	0.940	0.935	0.961	0.958
RF	0.866	0.831	0.859	0.845	0.858	0.834	0.839	0.834	0.847	0.867
MARS	0.913	0.874	0.897	0.845	0.846	0.902	0.932	0.839	0.902	0.885
MaxEnt	0.957	0.940	0.951	0.944	0.947	0.961	0.954	0.955	0.965	0.963

* GLM=Generalized Linear Model, GAM=Generalized Additive Model, ANN=Artificial Neural Network, SRE=Surface Range Envelope, CTA=Classification Tree Analysis, GBM=Generalised Boosted Model, RF=Random Forests, MARS=Multiple Adaptive Regression Splines, MaxEnt=Maximum Entropy

Table S3 Internal primers used for sequencing chloroplast *trn*S-*trn*G region

Primer	Sequence
trnS-IN-R	CAATTTTTTATATTCTATTATTATATAGAATTATAG
trnS-IN-F	CTATAATTCTATATAATAATAGAATATAAAAAATTG
trnS-IN-R1	AGAATTCTAAATATAATAGAATTAATAAAT
trnS-IN-F1	CGATATTTATTAATTCTATATTTAGAATTCT
trnS-IN-R2	AGTTATTAAGTTAACTATTTATTTCTATTTG
trnS-IN-F2	AGAATTGAAAATTGAAATATATAGAAATTCA

Supplementary Figures

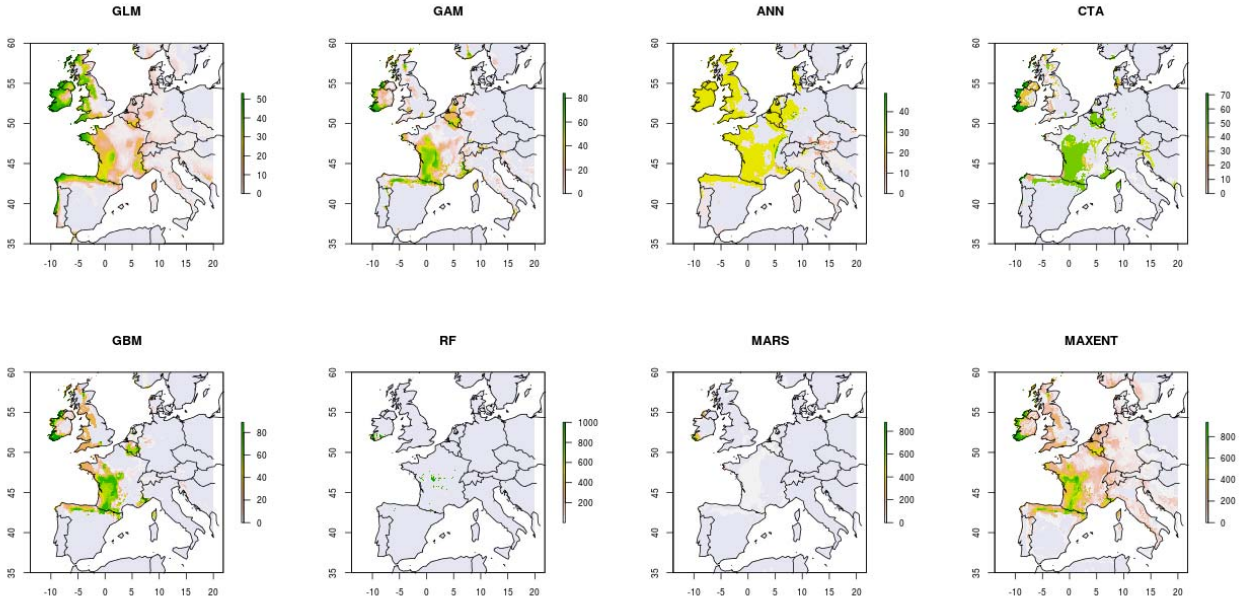


Figure S1. The eight model types used to generate the ensemble model showing the climate suitability for *Euphorbia hyberna* in the current climate. Here the suitability is shown with no range edge cut-off (compare with Fig.1 in the main text).

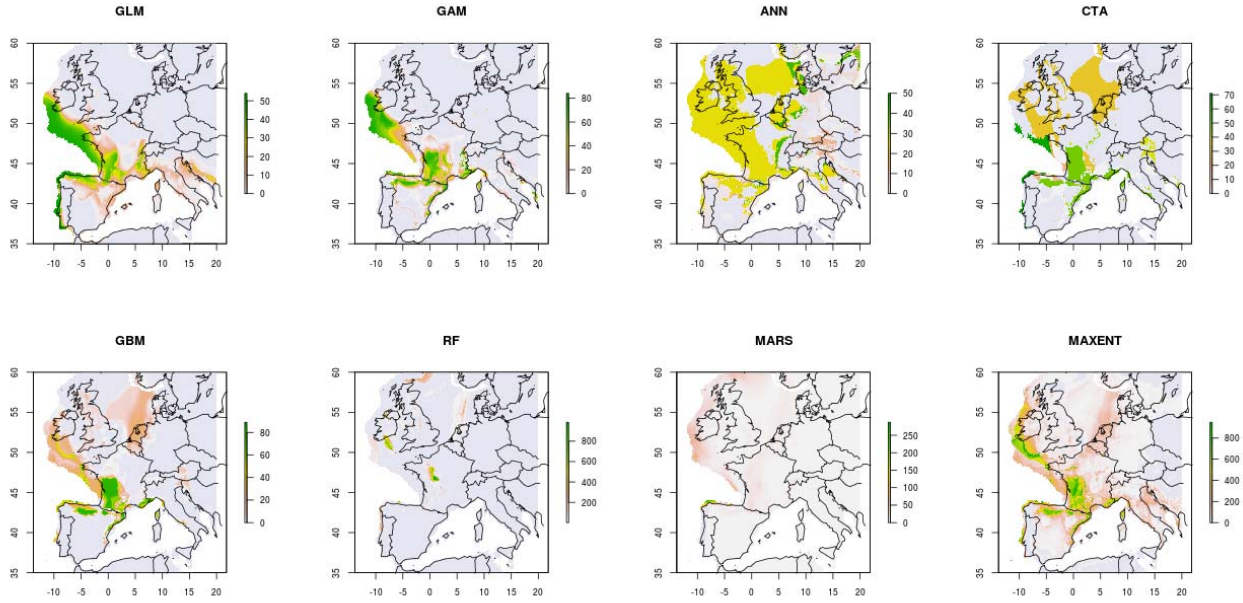


Figure S2. Similar to Fig. S1 but showing the modelled distribution of *Euphorbia hyberna* at the LGM.

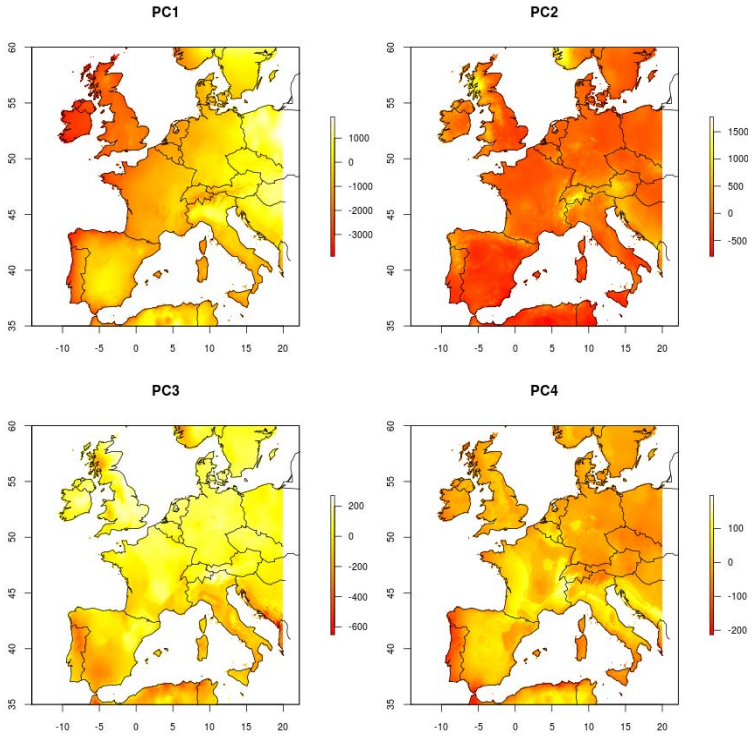


Figure S3. The four covariates used in the species distribution modelling for the current climate. These are the first four principal component scores of the 19 variable WorldClim dataset.

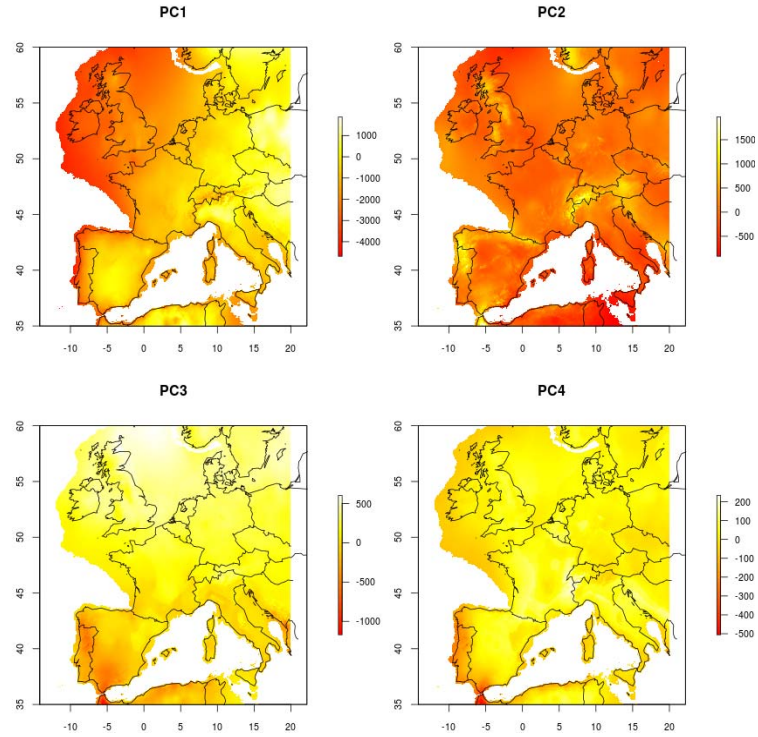


Figure S4. The four covariates used in the species distribution modelling for the LGM climate. These are PC scores based on the current climate PC but using the same variables from the CCSM climate dataset.

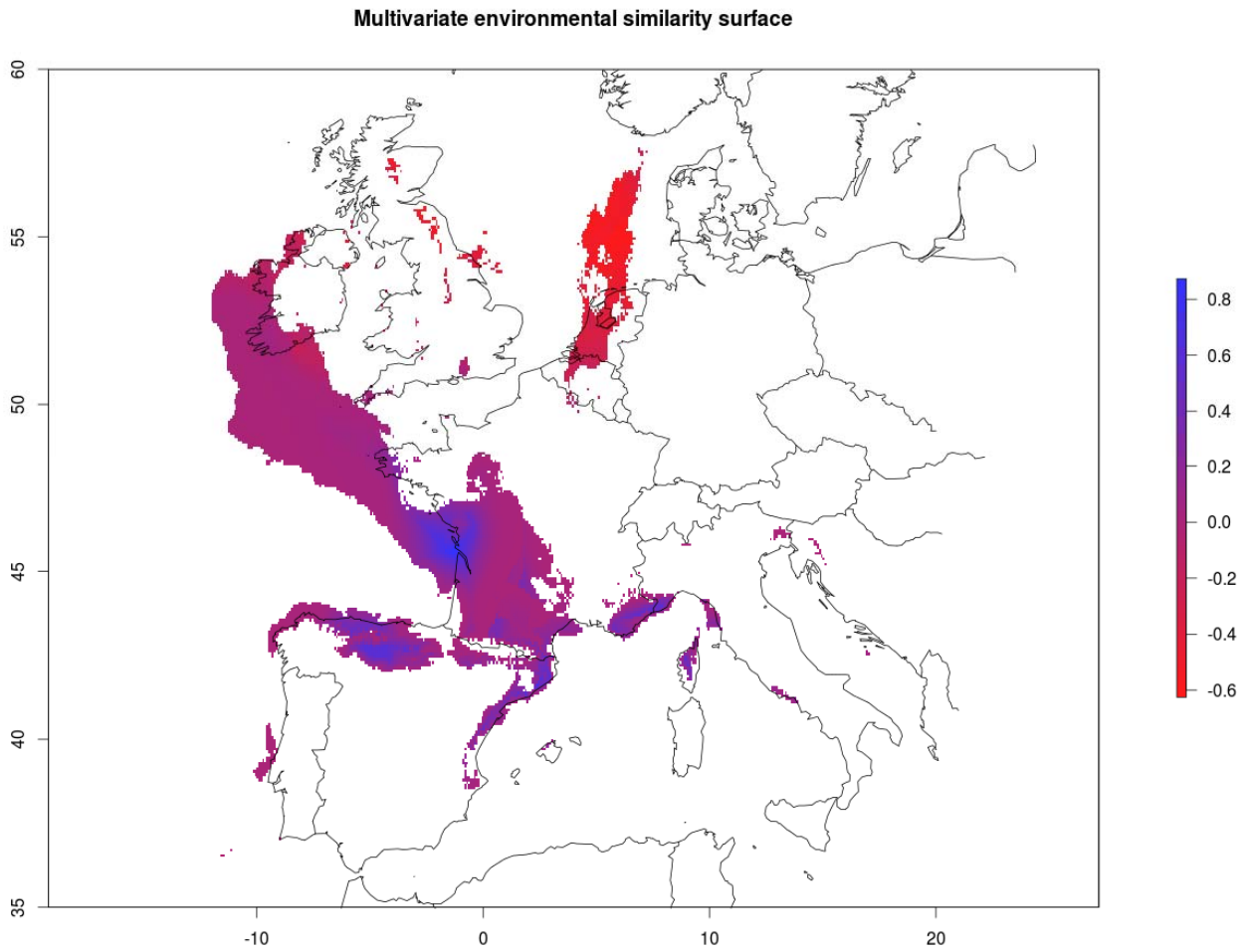


Figure S5. Multivariate environmental similarity surface (MESS) representing the similarity of the current climate, from which the species distribution models were constructed, to the LGM climate, for which they were projected. The colour scale represents similarity, such that low values (red) indicate a relatively large difference in climate while high values (blue) indicate small differences. Extrapolation into novel climate space is indicated by values below zero. Such areas only occur in the north, mainly where the ice sheet was present. The coloured area represents the projected distribution of Euphorbia hyberna at the LGM from the ensemble model. The MESS was calculated using the climate within the species range modelled in the current climate as the reference values compared with the climate within this LGM distribution (i.e. a comparison of climate within the two panels of Fig.2 in the main text).